

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/128432/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Zwanenburg, Alex, Vallières, Martin, Abdalah, Mahmoud A., Aerts, Hugo J. W. L., Andrearczyk, Vincent, Apte, Aditya, Ashrafinia, Saeed, Bakas, Spyridon, Beukinga, Roelof J., Boellaard, Ronald, Bogowicz, Marta, Boldrini, Luca, Buvat, Irène, Cook, Gary J. R., Davatzikos, Christos, Depeursinge, Adrien, Desseroit, Marie-Charlotte, Dinapoli, Nicola, Dinh, Cuong Viet, Echegaray, Sebastian, El Naqa, Issam, Fedorov, Andriy Y., Gatta, Roberto, Gillies, Robert J., Goh, Vicky, Götz, Michael, Guckenberger, Matthias, Ha, Sung Min, Hatt, Mathieu, Isensee, Fabian, Lambin, Philippe, Leger, Stefan, Leijenaar, Ralph T.H., Lenkiewicz, Jacopo, Lippert, Fiona, Losnegård, Are, Maier-Hein, Klaus H., Morin, Olivier, Müller, Henning, Napel, Sandy, Nioche, Christophe, Orlhac, Fanny, Pati, Sarthak, Pfaehler, Elisabeth A.G., Rahmim, Arman, Rao, Arvind U.K., Scherer, Jonas, Siddique, Muhammad Musib, Sijtsema, Nanna M., Fernandez, Jairo Socarras, Spezi, Emiliano ORCID: <https://orcid.org/0000-0002-1452-8813>, Steenbakkers, Roel J. H. M., Tanadini-Lang, Stephanie, Thorwarth, Daniela, Troost, Esther G. C., Upadhaya, Taman, Valentini, Vincenzo, van Dijk, Lisanne V., van Griethuysen, Joost, van Velden, Floris H. P., Whybra, Philip, Richter, Christian and Löck, Steffen 2020. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high throughput image-based phenotyping. *Radiology* 295 (2) , pp. 328-338. 10.1148/radiol.2020191145 file

Publishers page: <http://dx.doi.org/10.1148/radiol.2020191145>
<<http://dx.doi.org/10.1148/radiol.2020191145>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping

Alex Zwanenburg, PhD* • Martin Vallières, PhD* • Mahmoud A. Abdalah, PhD • Hugo J. W. L. Aerts, PhD • Vincent Andrearczyk, PhD • Aditya Apte, PhD • Saeed Ashrafinia, PhD • Spyridon Bakas, PhD • Roelof J. Beukinga, PhD • Ronald Boellaard, PhD • Marta Bogowicz, PhD • Luca Boldrini, PhD • Irène Buvat, PhD • Gary J. R. Cook, PhD • Christos Davatzikos, PhD • Adrien Depeursinge, PhD • Marie-Charlotte Desseroit, PhD • Nicola Dinapoli, PhD • Cuong Viet Dinh, PhD • Sebastian Echegaray, PhD • For the Group¹


From OncoRay—National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden-Rossendorf, Fetscherstr 74, PF 41, 01307 Dresden, Germany (A.Z., S. Leger, E.G.C.T., C.R., S. Lück); National Center for Tumor Diseases (NCT), Partner Site Dresden, Germany; German Cancer Research Center (DKFZ), Heidelberg, Germany, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Germany, and Helmholtz Association/Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Dresden, Germany (A.Z., S. Leger, E.G.C.T.); German Cancer Consortium (DKTK), Partner Site Dresden, and German Cancer Research Center (DKFZ), Heidelberg, Germany (A.Z., S. Leger, E.G.C.T., C.R., S. Lück); Medical Physics Unit, McGill University, Montréal, Canada (M.V., I.E.N.); Image Response Assessment Team Core Facility, Moffitt Cancer Center, Tampa, Fla (M.A.A.); Dana-Farber Cancer Institute, Brigham and Women's Hospital, and Harvard Medical School, Harvard University, Boston, Mass (H.J.W.L.A.); Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland (V.A., A.D., H.M.); Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY (A.A.); Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Md (S.A.); Department of Radiology and Radiological Science, Johns Hopkins University, Baltimore, Md (S.A., A.R.); Center for Biomedical Image Computing and Analytics (CBICA), University of Pennsylvania, Philadelphia, Pa (S.B., C.D., S.M.H., S.P.); Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pa (S.B., C.D., S.M.H., S.P.); Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pa (S.B.); Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen (UMCG), Groningen, the Netherlands (R.J.B., R.B., E.A.G.P.); Radiology and Nuclear Medicine, VU University Medical Centre (VUMC), Amsterdam, the Netherlands (R.B.); Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Zurich, Switzerland (M.B., M. Guckenberger, S.T.L.); Fondazione Policlinico Universitario "A. Gemelli" IRCCS, Rome, Italy (L.B., N.D., R.G., J.L., V.V.); Laboratoire d'Imagerie Translationnelle en Oncologie, Université Paris Saclay, Inserm, Institut Curie, Orsay, France (I.B., C.N., F.O.); Cancer Imaging Dept, School of Biomedical Engineering and Imaging Sciences, King's College London, London, United Kingdom (G.J.R.C., V.G., M.M.S.); Department of Nuclear Medicine and Molecular Imaging, Lausanne University Hospital, Lausanne, Switzerland (A.D.); Laboratory of Medical Information Processing (LaTIM)—team ACTION (image-guided therapeutic action in oncology), INSERM, UMR 1101, IBSAM, UBO, UBL, Brest, France (M.C.D., M.H., T.U.); Department of Radiation Oncology, the Netherlands Cancer Institute (NKI), Amsterdam, the Netherlands (C.V.D.); Department of Radiology, Stanford University School of Medicine, Stanford, Calif (S.E., S.N.). Received May 22, 2019; revision requested July 3; revision received December 9; accepted January 6, 2020. Address correspondence to A.Z. Fetscherstraße 74, PO Box PF 41, 01307 Dresden, Germany. (e-mail: alexander.zwanenburg@nct-dresden.de).

* A.Z. and M.V. contributed equally to this work.

¹ The complete list of authors and affiliations is at the end of this article.

Conflicts of interest are listed at the end of this article.

See also the editorial by Kuhl and Truhn in this issue.

Radiology 2020; 295:328–338 • <https://doi.org/10.1148/radiol.2020191145> • Content code: 

Background: Radiomic features may quantify characteristics present in medical imaging. However, the lack of standardized definitions and validated reference values have hampered clinical use.

Purpose: To standardize a set of 174 radiomic features.

Materials and Methods: Radiomic features were assessed in three phases. In phase I, 487 features were derived from the basic set of 174 features. Twenty-five research teams with unique radiomics software implementations computed feature values directly from a digital phantom, without any additional image processing. In phase II, 15 teams computed values for 1347 derived features using a CT image of a patient with lung cancer and predefined image processing configurations. In both phases, consensus among the teams on the validity of tentative reference values was measured through the frequency of the modal value and classified as follows: less than three matches, weak; three to five matches, moderate; six to nine matches, strong; 10 or more matches, very strong. In the final phase (phase III), a public data set of multimodality images (CT, fluorine 18 fluorodeoxyglucose PET, and T1-weighted MRI) from 51 patients with soft-tissue sarcoma was used to prospectively assess reproducibility of standardized features.

Results: Consensus on reference values was initially weak for 232 of 302 features (76.8%) at phase I and 703 of 1075 features (65.4%) at phase II. At the final iteration, weak consensus remained for only two of 487 features (0.4%) at phase I and 19 of 1347 features (1.4%) at phase II. Strong or better consensus was achieved for 463 of 487 features (95.1%) at phase I and 1220 of 1347 features (90.6%) at phase II. Overall, 169 of 174 features were standardized in the first two phases. In the final validation phase (phase III), most of the 169 standardized features could be excellently reproduced (166 with CT; 164 with PET; and 164 with MRI).

Conclusion: A set of 169 radiomics features was standardized, which enabled verification and calibration of different radiomics software.

© RSNA, 2020

Online supplemental material is available for this article.

Personalization of medicine is driven by the need to accurately diagnose disease and define suitable treatments for patients (1). Medical imaging is a potential source of biomarkers because it provides a macroscopic view of tissues of interest (2). Imaging has the advantage of being

noninvasive, readily available in clinical care, and repeatable (3,4).

Radiomics extracts features from medical imaging that quantify its phenotypic characteristics in an automated, high-throughput manner (5). Such features may help

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

Abbreviations

IBSI = Image Biomarker Standardization Initiative, ROI = region of interest

Summary

The Image Biomarker Standardization Initiative validated consensus-based reference values for 169 radiomics features, thus enabling calibration and verification of radiomics software.

Key Results

- Twenty-five research teams found agreement for calculation of 169 radiomics features derived from a digital phantom and a CT scan of a patient with lung cancer.
- Among these 169 standardized radiomics features, good to excellent reproducibility was achieved for 167 radiomics features using MRI, fluorine 18 fluorodeoxyglucose PET, and CT images obtained in 51 patients with soft-tissue sarcoma.

prognosticate, predict treatment outcomes, and assess tissue malignancy in cancer research (6–9). In neuroscience, features may help detect Alzheimer disease (10) and diagnose autism spectrum disorder (11).

Despite the growing clinical interest in radiomics, published studies have been difficult to reproduce and validate (5,9,12–14). Even for the same image, two different software implementations will often produce different feature values. This is because standardized definitions of radiomics features with verifiable reference values are lacking, and the image processing schemes required to compute features are not implemented consistently (15–18). This is exacerbated by reporting that is insufficiently detailed to enable studies and findings to be reproduced (19).

We formed the Image Biomarker Standardization Initiative (IBSI) to address these challenges by fulfilling the following objectives: (a) to establish nomenclature and definitions for commonly used radiomics features; (b) to establish a general radiomics image processing scheme for calculation of features from imaging; (c) to provide data sets and associated reference values for verification and calibration of software implementations for image processing and feature computation; and (d) to provide a set of reporting guidelines for studies involving radiomic analyses.

Materials and Methods

Study Design

We divided the current work into three phases (Fig 1). The first two phases focused on iterative standardization and were followed by a third validation phase. In phase I, the main objective was to standardize radiomics feature definitions and to define reference values, in the absence of any additional image processing. In phase II, we defined a general radiomics image processing scheme and obtained reference values for features under different image processing configurations. In phase III, we assessed if the standardization conducted in the previous phases resulted in reproducible feature values for a validation data set.

Research Teams

We invited teams of radiomics researchers to collaborate in the IBSI. Participation was voluntary and open for the duration of

the study. Teams were eligible if they (a) developed their own software for image processing and feature computation and (b) could participate in any phase of the study.

Radiomics Features

We defined a set of 174 radiomics features (Table 1). This set consisted of features commonly used to quantify the morphologic characteristics, first-order statistical aspects, and spatial relationships between voxels (texture) in regions of interest (ROIs) in three-dimensional images. To compute texture features, additional feature-specific parameters were required. This increased the number of computed features beyond 174 (Appendix E1 [online]). All feature definitions are provided in chapter 3 of the IBSI reference manual (Appendix E2 [online]).

General Radiomics Image Processing Scheme

We defined a general radiomics image processing scheme based on descriptions in the literature (3,6,17,20). The scheme contained the main processing steps required for computation of features from a reconstructed image and is depicted in Figure 2. A full description of these image processing steps may be found in chapter 2 of the IBSI reference manual (online).

Data Sets

Each phase used a different data set. In phase I, we designed a small 80-voxel three-dimensional digital phantom with a 74-voxel ROI mask to facilitate the process of establishing reference values for features, without involving image processing.

In phase II, we used a publicly available CT image in a patient with lung cancer. The accompanying segmentation of the gross tumor volume was used as the ROI (21).

The validation data set used in phase III consisted of a cohort of 51 patients with soft-tissue sarcoma who underwent multi-modality imaging (coregistered CT, fluorine 18 fluorodeoxyglucose PET, and T1-weighted MRI) from the Cancer Imaging Archive (20,22,23). Each image was accompanied by a gross tumor volume segmentation, which was used as the ROI. PET and MRI were centrally preprocessed (Appendix E1 [online]) to ensure that standardized uptake value conversion and bias-field correction steps did not affect validation.

Defining Consensus on the Validity of Feature Reference Values

In the first two phases, research teams computed feature values from the ROI in the associated image data set directly (phase I) and according to predefined image processing parameters (phase II; Appendix E1 [online]). All of the most recent values submitted by each team were collected and limited to three significant digits. Then, we used the mode of the submitted values for each feature as a tentative reference value.

We quantified the level of consensus on the validity of a tentative reference value for each feature using two measures: (a) the number of research teams that submitted a value that matched the tentative reference value within a tolerance margin (Appendix E1 [online]) and (b) the previous number divided by the total number of research teams that submitted a value.

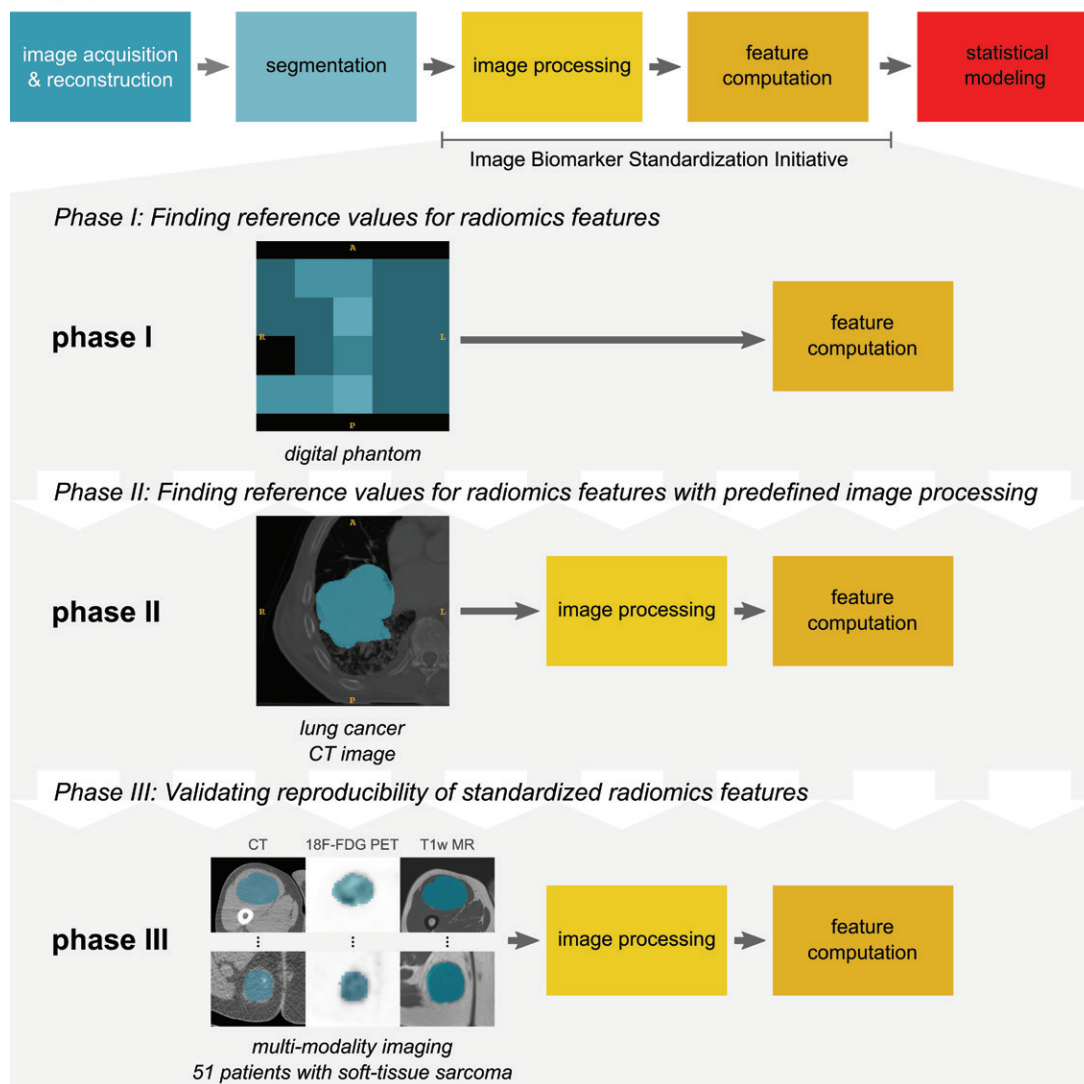
Imaging and radiomics workflow

Figure 1: Flowchart of study overview. The workflow in a typical radiomics analysis starts with acquisition and reconstruction of a medical image. Subsequently, the image is segmented to define regions of interest (ROI). Afterward, radiomics software is used to process the image and to compute features that characterize an ROI. We focused on standardizing the image processing and feature computation steps. Standardization was performed within two iterative phases. In phase I, we used a specially designed digital phantom to obtain reference values for radiomics features directly. In phase II, a publicly available CT image in a patient with lung cancer was used to obtain reference values for features under predefined configurations of a standardized general radiomics image processing scheme. Standardization of image processing and feature computation steps in radiomics software was prospectively validated during phase III by assessing reproducibility of standardized features in a publicly available multimodality patient cohort of 51 patients with soft-tissue sarcoma. 18F-FDG = fluorine 18 fluorodeoxyglucose, T1w = T1-weighted.

Four consensus levels were assigned based on the first consensus measure as follows: less than three, weak; three to five, moderate; six to nine, strong; 10 or more, very strong. The second measure assessed the stability of the consensus. We considered a tentative reference value for a feature to be valid only if it had at least moderate consensus and it was reproduced by an absolute majority (exceeding 50%) of the contributing research teams.

Iterative Standardization Process

In the first two phases, we iteratively refined consensus on the validity of feature reference values. This iterative process simultaneously served to standardize feature definitions and the gen-

eral radiomics image processing scheme (24). At the start of the iterative process, we provided initial definitions for features (phase I) and the general radiomics image processing scheme (phase II) in a working document. For phase I, we manually calculated mathematically exact reference values for all but morphologic features to verify values produced by the research teams. For phase II, we defined five different image processing configurations (configurations A–E) that covered a range of image processing parameters and methods commonly used in radiomics studies (Appendix E1 [online]).

After producing the initial working document, we asked the research teams to compute feature values from the ROI in the

Table 1: Overview of Included Radiomics Features

Feature Family	Base Definition	No. of Features			
		Phase I	Phase II Configuration A–B (2D)	Phase II Configuration C–E (3D)	Phase III
Morphologic characteristics	29	29	29	29	29
Local intensity	2	2	2	2	2
Intensity-based statistics	18	18	18	18	18
Intensity histogram	23	23	23	23	23
Intensity-volume histogram	7	7	7	7	7
Gray-level co-occurrence matrix*	25	150	100	50	25
Gray-level run-length matrix*	16	96	64	32	16
Gray-level size-zone matrix*	16	48	32	16	16
Gray-level distance-zone matrix*	16	48	32	16	16
Neighborhood gray tone difference matrix*	5	15	10	5	5
Neighboring gray level dependence matrix*	17	51	34	17	17
Total	174	487	351	215	174

Note.—Data are numbers of features. A set of 174 radiomics features was standardized and validated in three phases. In phase I, features were computed without any prior image processing. In phase II, features were computed after image processing with five predefined configurations (configurations A–E; image slice-based (two-dimensional [2D]) or volumetric (three-dimensional [3D]); Appendix E1 [online]). In the final phase III, we assessed the reproducibility of features standardized in phases I and II.

* Texture features have additional parameters that are required for their calculation, which increased the number of computed features (Appendix E1 [online]).

digital phantom (phase I) and from the ROI in the lung cancer CT image after image processing according to the different predefined image processing configurations (phase II). Feature values were collected and processed to analyze the consensus on the validity of tentative reference values. The results were then made available to all teams at an average interval of 4 weeks. The study leader would also contact the teams with feedback after comparing their submitted feature values with the mathematically exact values (phase I only) and with feature values obtained by other teams (phases I and II). The research teams provided feedback in the form of questions and suggestions concerning the descriptions in the working document and the standardization of radiomics software. The working document was regularly updated as a result. Teams would then make changes to their software based on the results of the analysis and feedback from the study leader.

The two iterative phases were staggered to make it easier to separate differences and errors related to feature computation from those related to image processing. The initial contributions from phase I were analyzed in September 2016. We initiated phase II after achieving moderate or better consensus on the validity of reference values for at least 70% of the features, that is, time point 6 (January 2017). Initial contributions for phase II were analyzed at time point 10 (April 2017). Afterward, phases I and II were concurrent. We halted the iterative standardization process at time point 25 (March 2019) after we attained strong or better consensus on validity of reference values for more than 90% of the features in both phases I and II. The overall timeline of the study is summarized in Appendix E1 (online).

Validation

After the standardization process finished, we asked the research teams to compute 174 features from the gross tumor

volume in each of the images in the soft-tissue sarcoma validation cohort using a realistic, predefined image processing configuration (Appendix E1, [online]). The computed feature values were collected and processed centrally, as follows. First, for each team, we removed any feature that was not standardized by their software. To do so, we compared the reference values of the respective feature with the values that the team obtained from the CT image in the patient with lung cancer under image processing configurations C, D, and E (as in phase II). If a value did not match its reference value, the feature was not used. The reproducibility of remaining standardized features was subsequently assessed using a two-way, random-effects, single-rater, absolute agreement intraclass correlation coefficient (ICC) (25). Using the lower boundary of the 95% confidence interval of the ICC value (26), reproducibility of each feature was assigned to one of the following categories, as suggested by Koo and Li (27): poor, lower boundary less than 0.50; moderate, lower boundary greater than or equal to 0.50 and less than 0.75; good: lower boundary greater than 0.75 and less than 0.90; excellent, lower boundary greater than 0.90.

Results

Characteristics of the Participating Research Teams

Twenty-five teams contributed to the IBSI (Fig 3; Appendix E1 [online]). Fifteen teams contributed to both standardization phases, and nine teams contributed to the validation phase. One team retired because they switched to software developed by another team. Five teams implemented 95% or more of the defined features. Nine teams were able to compute features for all image processing configurations in phase II (Appendix E1 [online]).

The University Medical Center Groningen and the French National Institute of Health and Medical Research provided

three and two teams of researchers, respectively. This did not compromise consensus on the validity of feature reference values. Moderate, strong, or very strong consensus on the validity of the reference values was based on teams from at least three, five, and eight different top-level institutions, respectively (Appendix E1 [online]).

Matlab ($n = 10$), C++ ($n = 7$), and Python ($n = 5$) were the most popular programming languages. No language dependency was found; consensus of all features with a moderate or better consensus on the validity of their reference values was based on multiple programming languages (Appendix E1 [online]).

Consensus on Validity of Feature Reference Values

Consensus on the validity of feature reference values improved during the course of the study, as shown in Figure 4 and Table 2. Initially, only weak consensus existed for the majority of features: 232 of 302 (76.8%) and 703 of 1075 (65.4%) for phase I and II, respectively.

At the final analysis time point, the number of features with a weak consensus had decreased to two of 487 (0.4%) for phase I and 19 of 1347 (1.4%) for phase II. The remaining features with weak consensus on the validity of their (tentative) reference values were the area and volume densities of the oriented minimum bounding box and the minimum volume enclosing ellipsoid (Appendix E1 [online]). We were unable to standardize the complex algorithms that are required to compute the oriented minimum bounding box and minimum volume enclosing ellipsoid. Therefore, the previous features should not be regarded as standardized.

As shown in Table 2, strong or better consensus could be established for 463 of 487 (95.1%) and 1220 of 1347 (90.6%) features in phases I and II, respectively. None of these features were found to be unstable. In phase II, two of 108 (1.9%)

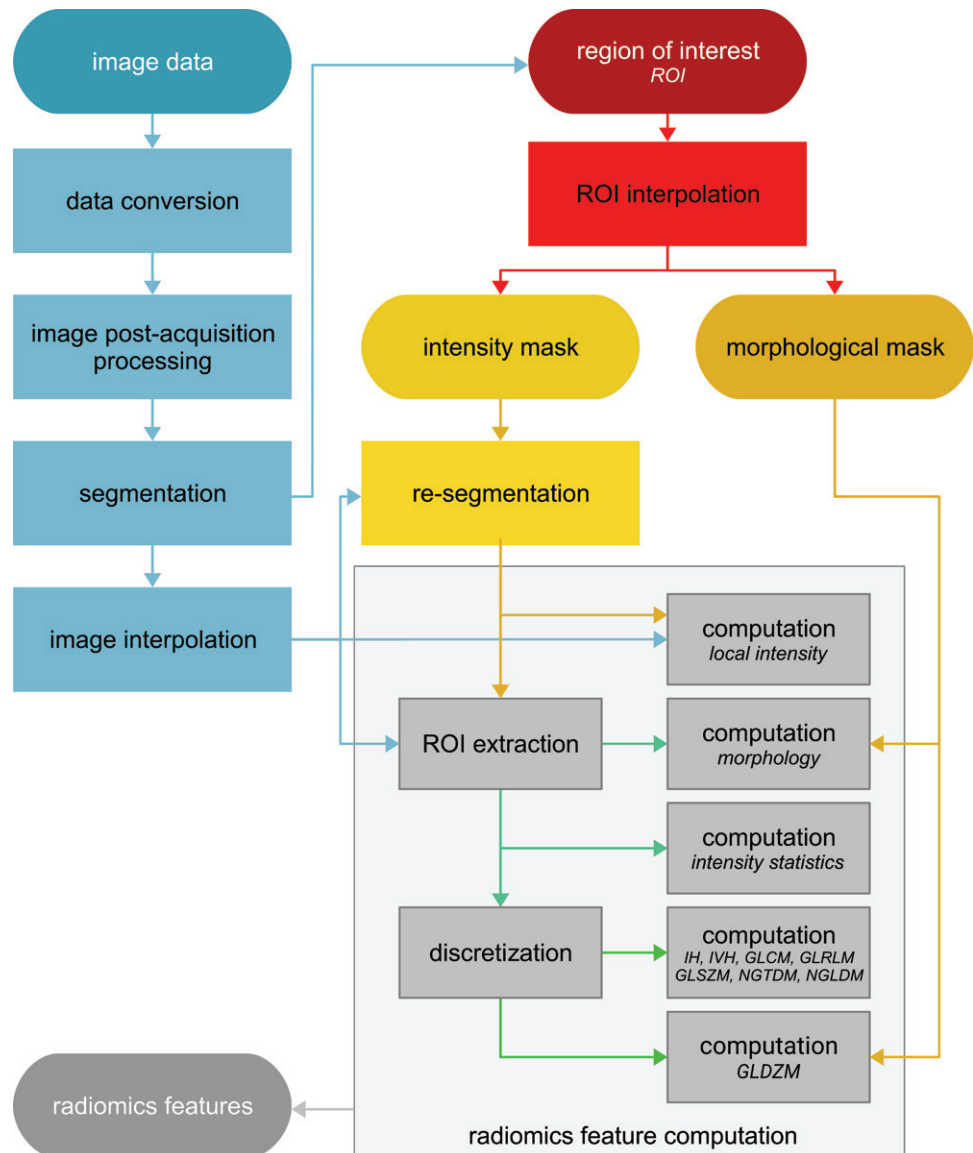


Figure 2: Flowchart of the general radiomics image processing scheme for computing radiomics features. Image processing starts with reconstructed images. These images are processed through several optional steps: data conversion (eg, conversion to standardized uptake values), image postacquisition processing (eg, image denoising), and image interpolation. Either the region of interest (ROI) is created automatically during the segmentation step, or an existing ROI is retrieved. The ROI is then interpolated as well, and intensity and morphologic masks are created as copies. The intensity mask may be resegmented according to intensity values to improve comparability of intensity ranges across a cohort. Radiomics features are then computed from the image masked by the ROI and its immediate neighborhood (local intensity features) or the ROI itself (all others). Image intensities are moreover discretized prior to computation of features from the intensity histogram (IH), intensity-volume histogram (IVH), gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level distance-zone matrix (GLDZM), neighborhood gray-tone difference matrix (NGTDM), and neighboring gray-level dependence matrix (NGLDM) families. All processing steps from image interpolation to the computation of radiomics features were evaluated in this study.

features with moderate consensus were unstable. Both were derived from the same feature: the area under the curve of the intensity-volume histogram. Hence, we do not consider this feature to be standardized.

The most commonly implemented features were mean, skewness, excess kurtosis, and minimum of the intensity-based statistics family. These were implemented by 23 of 24 research teams. No feature was implemented by all teams (Appendix E1 [online]).

Reproducibility of Standardized Features

We were able to find stable reference values with moderate or better consensus for 169 of 174 features. In the validation phase, most of these features could be reproduced well (Fig 5, Appendix E1 [online]). Excellent reproducibility was found for 166 of 174, 164 of 174, and 164 of 174 features for CT, PET, and MRI, respectively. Good reproducibility was found for one of 174 (CT) and three of 174 (PET and MRI) features. For each modality, two of 174 features had unknown reproducibility, which indicated that they were computed by fewer than two teams during validation. These features were Moran's I index and Geary's C measure. Although they were standardized, they were expensive to compute. The remaining five of 174 features could not be standardized during the first two phases and were not assessed during validation.

Discussion

In this study, the Image Biomarker Standardization Initiative (IBSI) produced and validated a set of consensus-based reference values for radiomics features. Twenty-five research teams were able to standardize 169 of 174 features, which were subsequently shown to have good to excellent reproducibility in a validation data set.

With the completion of the current work, compliance with the IBSI standard can be checked for any radiomics software, as follows.

First, use the software to compute features using the digital phantom. Compare the resulting values against the reference values found in the IBSI reference manual and the compliance check spreadsheet created for this purpose (Appendix E3 [online]). Investigate any difference. Subsequently, resolve the differences or explain them (eg, the use of kurtosis instead of excess kurtosis).

Then repeat the previous steps with the CT data set used in this study and one or more of the image processing configurations used in phase II.

Initial consensus on the validity of reference values for many features was weak, which means that teams obtained different values for the same feature. This mirrored findings reported elsewhere (15–18). Several notable causes of deviations were

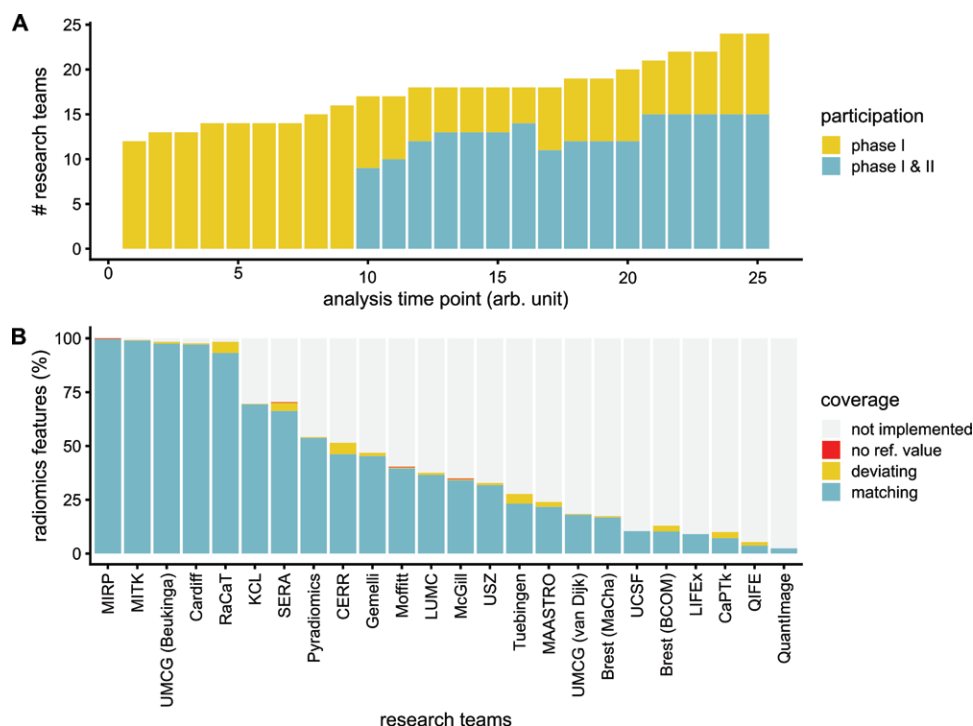


Figure 3: Bar graphs depict participation and radiomics feature coverage by research teams. A, Graph shows the number of research teams at each analysis time point during the two phases of the iterative standardization process. Teams computed features without prior image processing (phase I) and after image processing (phase II), with the aim of finding reference values for a feature. Consensus on the validity of reference values was assessed at each time point, the time between which was variable (arbitrary unit [arb. unit]). B, Graph shows the final coverage of radiomics features implemented by each team in phase I, as well as the team's ability to reproduce the reference value of a feature. We were unable to obtain reliable reference values for five features (no ref. value). The teams are listed in Appendix E1 (online). BCOM = Institute of Research and Technology bcom, Brest; CaPTk = Cancer Imaging Phenomics Toolkit; CERR = Computational Environment for Radiological Research; KCL = King's College London; LUMC = Leiden University Medical Center; MAASTRO = Maastricht, Maastricht, the Netherlands; MaCha = Marie-Charlotte Desseroit; MIRP = Medical Image Radiomics Processor; MITK = Medical Imaging Interaction Toolkit; QIFE = Quantitative Image Feature Engine; RaCaT = Radiomics Calculator; SERA = Standardized Environment for Radiomics Analysis; UCSF = University of California, San Francisco; UMCG = University Medical Center Groningen; USZ = University of Zurich.

identified—for example, differences in interpolation, morphologic representation of the ROI, and nomenclature differences—and were subsequently resolved (Appendix E1 [online]). In effect, we cross-calibrated radiomics software implementations.

The demonstrated lack of initial correspondence between teams carries a clinical implication. Software implementations of seemingly well-defined mathematic formulas can vary greatly in the numeric results they produce. Clinical radiologists who are using advanced image analysis workstations should be aware of this, think critically about comparing results produced by different workstations, and demand more details and validation studies from the vendors of those workstations.

Findings from most radiomics studies have not been translated into clinical practice, and they require external retrospective and prospective validation in clinical trials (2,28). The IBSI, in addition to the presented work, has defined reporting guidelines that indicate the elements that should be reported to facilitate this process. However, we refrained from creating a comprehensive recommendation on how to perform a good radiomics analysis for several reasons. First, such recommendations will need to be modality specific and possibly entity specific (29,30). The related specific evidence for the effect of particular

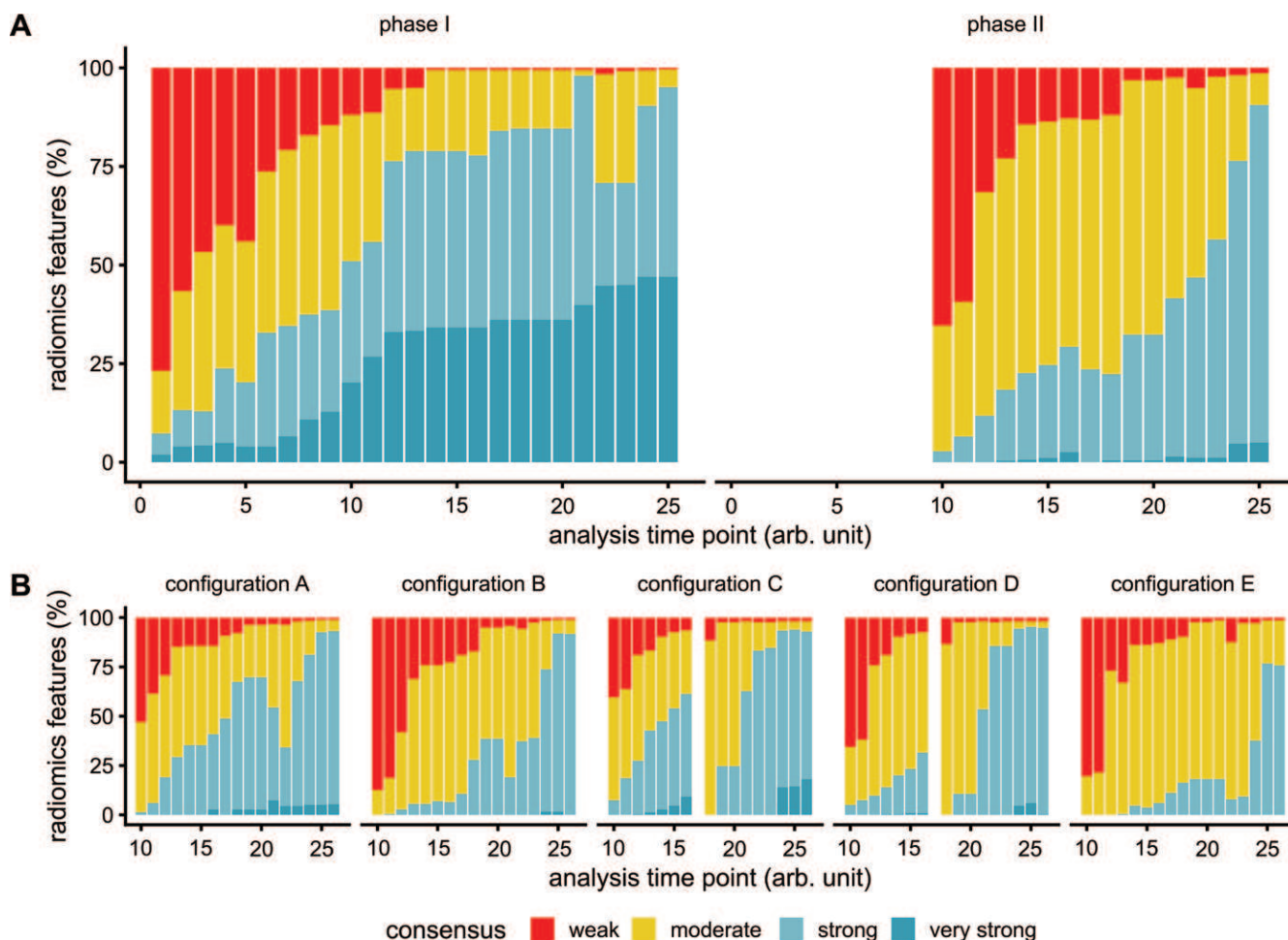


Figure 4: Bar graphs depict iterative development of consensus on the validity of reference values for radiomics features. We tried to find reliable reference values for radiomics features in an iterative standardization process. In phase I, features were computed without prior image processing, whereas in phase II, features were assessed after image processing with five predefined configurations (configurations A–E; Appendix E1 [online]). The panels show, A, the overall development of consensus on the validity of (tentative) reference values in phases I and II and, B, the development of consensus in phase II, according to image processing configuration. Consensus on the validity of a reference value is based on the number of research teams that produce the same value for a feature (weak: ≤ 3 ; moderate: three to five; strong: six to nine; very strong: ≥ 10). We analyzed consensus at each of the analysis time points, the time between which was variable (arbitrary unit; arb. unit). New features were included at time points 5 and 22, causing an apparent decrease in consensus. For phase II, we first analyzed consensus at time point 10. Image processing configurations C and D were altered after time point 16. Configuration E was altered after revising the resegmentation processing step at time point 22. See Appendix E1 (online) for more information regarding the timeline.

parameters, for example, the choice of interpolation algorithm, is far from complete. Second, recommendations or guidelines regarding parts of the radiomics analysis are already covered comprehensively elsewhere, for example, by the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis statement on diagnostic and prognostic modeling (31). Certainly, the image processing configurations used in phase II are not intended for general use, as their primary aim was to cover a range of different methods. Only the configurations defined for the validation data set resemble a realistic set of parameters given the entity and imaging modalities.

Our study has several limitations. First, our aim was to lay a foundation for standardized computation of radiomics features. To this end, we sought to standardize 174 commonly used features and to obtain reference values using image processing methods that radiomics researchers most commonly employ. To keep the scope manageable, many other features such as fractals

and image filters were not assessed (32), important modality-specific image processing steps were not benchmarked, and uncommon image processing methods were not investigated. This is a serious limitation and one that the IBSI is currently addressing.

Despite the fact that standardized feature computation is an important step toward reproducible radiomics, the need for standardization and harmonization related to image acquisition, reconstruction, and segmentation remains, as these constitute additional sources of variability in radiomics studies. Because of this variability, features that can be reproduced from the same image using standardized radiomics software may nevertheless lack reproducibility in multicentric or multiscanner settings (14,19,33). We did not address these issues here as their comprehensive harmonization is the ongoing focus of other consortia and professional societies (2). Other approaches have also been proposed to address these issues, such as the reduction of cohort effects on radiomics features using statistical methods (34) and

Table 2: Consensus on the Validity of Reference Values of Radiomics Features at Initial and Final Analysis Time Points for Phases I and II

Parameter	Total		Weak Consensus		Moderate Consensus		Strong Consensus		No. of Features with Very Strong Consensus	No. of Features with Moderate Consensus or Greater	No. of Features with Strong Consensus or Greater
	No. of Features	No. of Unstable Features	No. of Features	No. of Unstable Features	No. of Features	No. of Unstable Features	No. of Features	No. of Unstable Features			
Initial analysis time point											
Phase I	302	147 (48.7)	232 (76.8)	133 (57.3)	48 (15.9)	12 (25.0)	16 (5.3)	2 (12.5)	6 (2.0)	70 (23.2)	22 (7.3)
Phase II	1075	610 (56.7)	703 (65.4)	537 (76.4)	342 (31.8)	73 (21.3)	30 (2.8)	0 (0)	0 (0)	372 (34.6)	30 (2.8)
Config A	215	28 (13.0)	114 (53.0)	26 (22.8)	98 (45.6)	2 (2.0)	3 (1.4)	0 (0)	0 (0)	101 (47.0)	3 (1.4)
Config B	215	149 (69.3)	188 (87.4)	149 (79.3)	27 (12.6)	0 (0)	0 (0)	0 (0)	0 (0)	27 (12.6)	0 (0)
Config C	215	97 (45.1)	87 (40.5)	72 (82.8)	112 (52.1)	25 (22.3)	16 (7.4)	0 (0)	0 (0)	128 (59.5)	16 (7.4)
Config D	215	162 (75.3)	141 (65.6)	129 (91.5)	63 (29.3)	33 (52.4)	11 (5.1)	0 (0)	0 (0)	74 (34.4)	11 (5.1)
Config E	215	174 (80.9)	173 (80.5)	161 (93.1)	42 (19.5)	13 (31.0)	0 (0)	0 (0)	0 (0)	42 (19.5)	0 (0)
Final analysis time point											
Phase I	487	2 (0.4)	2 (0.4)	2 (100.0)	22 (4.5)	0 (0)	234 (48.0)	0 (0)	229 (47.0)	485 (99.6)	463 (95.1)
Phase II	1347	20 (1.5)	19 (1.4)	18 (94.7)	108 (8.0)	2 (1.9)	1152 (85.5)	0 (0)	68 (5.0)	1328 (98.6)	1220 (90.6)
Config A	351	4 (1.1)	4 (1.1)	3 (75.0)	22 (6.3)	1 (4.5)	307 (87.5)	0 (0)	18 (5.1)	347 (98.9)	325 (92.6)
Config B	351	5 (1.4)	4 (1.1)	4 (100.0)	24 (6.8)	1 (4.2)	317 (90.3)	0 (0)	6 (1.7)	347 (98.9)	323 (92.0)
Config C	215	4 (1.9)	4 (1.9)	4 (100.0)	9 (4.2)	0 (0)	171 (79.5)	0 (0)	31 (14.4)	211 (98.1)	202 (94.0)
Config D	215	4 (1.9)	4 (1.9)	4 (100.0)	6 (2.8)	0 (0)	192 (89.3)	0 (0)	13 (6.0)	211 (98.1)	205 (95.3)
Config E	215	3 (1.4)	3 (1.4)	3 (100.0)	47 (21.9)	0 (0)	165 (76.7)	0 (0)	0 (0)	212 (98.6)	165 (76.7)

Note.—Data are number of features, with percentages in parentheses. Reference values of radiomics features were iteratively obtained in two phases. In phase I, features were computed without prior image processing, whereas in phase II, features were computed after image processing with five predefined configurations (configurations A–E; Appendix E1 [online]). Consensus on the validity of a reference value was based on the number of research teams that produced the same value (weak: ≤ 3 ; moderate: three to five; strong: six to nine; very strong: ≥ 10). Unstable features are those for which the consensus was only carried by a minority of teams ($\leq 50\%$). Features with very strong consensus were never unstable, and the respective column was omitted. The number of features increased between the initial and final time points due to adding new features and computing features with additional feature-specific parameters (Appendix E1 [online]). Config = configuration.

application of artificial intelligence to convert between reconstruction kernels in CT imaging (35).

In conclusion, the Image Biomarker Standardization Initiative was able to produce and validate reference values for radiomics features. These reference values enable verification of radiomics software, which will increase reproducibility of radiomics studies and facilitate clinical translation of radiomics.

Acknowledgments: The authors thank Baptiste Laurent, MSc, Sarah Mattonen, PhD, Hesham Elhalawani, PhD, Jayashree Kalpathy-Cramer, PhD, Dennis Mackin, PhD, Ida A. Nissen, PhD, Dimitris Visvikis, PhD, and Maqsood Yaqub, PhD, for

their valuable ideas and support. In addition, we thank Rutu Pandya, MSc, and Roger Schaefer, BSc, for technical support in setting up and administrating the IBSI website (<https://theibsi.github.io/>). We also thank David Clunie, MBBS, for his input on creating permanent IBSI identifiers and providing a Digital Imaging and Communications in Medicine version of the digital phantom. We thank Alberto Traverso, PhD, for integrating the work of the IBSI in the Radiomics Ontology (<https://biportal.bioontology.org/ontologies/RO/>). Special thanks to the European Society for Radiation Therapy and Oncology and to Uulke van der Heide, PhD, for organizing a Radiomics Mini Workshop where the idea for a standardization initiative was first discussed.

Complete list of authors: Alex Zwanenburg, PhD; Martin Vallières, PhD; Mahmoud A. Abdallah, PhD; Hugo J.W. L. Aerts, PhD; Vincent Andrearczyk, PhD; Aditya Apte, PhD; Saeed Ashrafinia, PhD; Spyridon Bakas, PhD; Roelof J. Beukinga, PhD; Ronald Boellaard, PhD; Marta Bogowicz, PhD; Luca Boldrini, PhD; Irène Buvat,

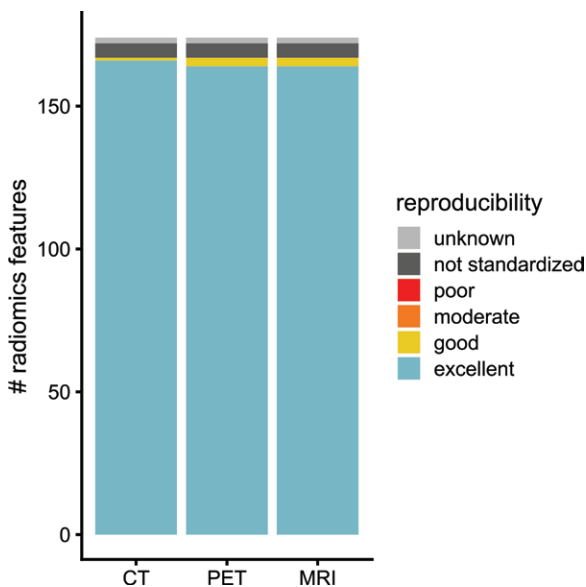


Figure 5: Bar graph shows reproducibility of standardized radiomics features. We assessed reproducibility of 169 standardized features on a validation cohort of 51 patients with soft-tissue sarcoma using multimodality imaging (CT, fluorine 18 fluorodeoxyglucose PET, and T1-weighted MRI; shown as CT, PET and MRI) according to the feature values computed by research teams. We assigned each feature to a reproducibility category based on the lower boundary of the 95% confidence interval of the two-way random effects, single rater, absolute agreement intraclass correlation coefficient of the feature (poor: <0.50 ; moderate: $0.50-0.75$; good: $0.75-0.90$; excellent: ≥ 0.90). Five features could not be standardized in this study. Two features with unknown reproducibility were computed by fewer than two teams during validation.

PhD; Gary J. R. Cook, PhD; Christos Davatzikos, PhD; Adrien Depeursinge, PhD; Marie-Charlotte Desseroit, PhD; Nicola Dinapoli, PhD; Cuong Viet Dinh, PhD; Sebastian Echegaray, PhD; Issam El Naqa, PhD; Andriy Y. Fedorov, PhD; Roberto Gatta, PhD; Robert J. Gillies, PhD; Vicky Goh, PhD; Michael Götz, PhD; Matthias Guckenberger, PhD; Sung Min Ha, MSc; Mathieu Hatt, PhD; Fabian Isensee, MSc; Philippe Lambin, PhD; Stefan Leger, PhD; Ralph T.H. Leijenaar, PhD; Jacopo Lenkowicz, MSc; Fiona Lippert, MSc; Are Losnegård, PhD; Klaus H. Maier-Hein, PhD; Olivier Morin, PhD; Henning Müller, PhD; Sandy Napel, PhD; Christophe Nioche, PhD; Fanny Orlhac, PhD; Sarthak Pati, MSc; Elisabeth A.G. Pfahler, MSc; Arman Rahmim, PhD; Arvind U.K. Rao, PhD; Jonas Scherer, MSc; Muhammad Musib Siddique, PhD; Nanna M. Sijtsema, PhD; Jairo Socarras Fernandez, MSc; Emiliano Spezi, PhD; Roel J.H.M. Steenbakkers, PhD; Stephanie Tanadini-Lang, PhD; Daniela Thorwarth, PhD; Esther G.C. Troost, PhD; Taman Upadhaya, PhD; Vincenzo Valentini, PhD; Lisanne V. van Dijk, PhD; Joost van Griethuysen, PhD; Floris H.P. van Velden, PhD; Philip Whybra, MSc; Christian Richter, PhD; Steffen Löck, PhD.

Author affiliations continued: Department of Radiation Oncology, Physics Division, University of Michigan, Ann Arbor, Mich (I.E.N., A.U.K.R.); Surgical Planning Laboratory, Brigham and Women's Hospital and Harvard Medical School, Harvard University, Boston, Mass (A.Y.F.); Department of Cancer Imaging and Metabolism, Moffitt Cancer Center, Tampa, Fla (R.J.G.); Department of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany (M. Götz, F.L., K.H.M.H., J.S.); The D-Lab, Department of Precision Medicine, GROW-School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, the Netherlands (P.L., R.T.H.L.); Section for Biomedical Physics, Department of Radiation Oncology, University of Tübingen, Germany (F.L., J.S.F., D.T.); Department of Clinical Medicine, University of Bergen, Bergen, Norway (A.L.); Department of Radiation Oncology, University of California, San Francisco, Calif (O.M.); University of Geneva, Geneva, Switzerland (H.M.); Department of Electrical Engineering, Stanford University, Stanford, Calif (S.N.); Department of Medicine (Biomedical Informatics Research), Stanford University School of Medicine, Stanford, Calif (S.N.); Departments of Radiology and Physics, University of British Columbia, Vancouver, Canada (A.R.); Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Mich (A.U.K.R.); Department of Radiation Oncology, University of Groningen, University Medical Center Groningen (UMCG), Groningen, the Netherlands

(N.M.S., R.J.H.M.S., L.V.v.D.); School of Engineering, Cardiff University, Cardiff, United Kingdom (E.S., P.W.); Department of Medical Physics, Velindre Cancer Centre, Cardiff, United Kingdom (E.S.); Department of Radiotherapy and Radiation Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany (E.G.C.T., C.R., S. Löck), Helmholtz-Zentrum Dresden-Rossendorf, Institute of Radiooncology—OncoRay, Dresden, Germany (E.G.C.T., C.R.); Department of Nuclear Medicine, CHU Milétrie, Poitiers, France (T.U.); Department of Radiology, the Netherlands Cancer Institute (NKI), Amsterdam, the Netherlands (J.v.G.); GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, the Netherlands (J.v.G.); Department of Radiation Oncology, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, Mass (J.v.G.); and Department of Radiology, Leiden University Medical Center (LUMC), Leiden, the Netherlands (F.H.P.v.V.).

Author contributions: Guarantors of integrity of entire study, A.Z., M.A.A., A.D., I.E.N., P.L., O.M., S.N., S.P., E.A.G.P., J.S.F., E.S., R.J.H.M.S., T.U., L.V.v.D.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, A.Z., M.V., M.A.A., H.J.W.L.A., G.J.R.C., C.D., N.D., C.V.D., I.E.N., A.Y.F., M.H., O.M., H.M., S.P., A.U.K.R., M.M.S., J.S.F., R.J.H.M.S., D.T., E.G.C.T., V.V., F.H.P.v.V., C.R.; clinical studies, G.J.R.C., N.D., I.E.N., A.Y.F., S.P., J.S.F., R.J.H.M.S., V.V., F.H.P.v.V.; experimental studies, A.Z., M.V., M.A.A., H.J.W.L.A., V.A., A.A., S.A., S.B., R.J.B., R.B., M.B., L.B., G.J.R.C., C.D., A.D., M.C.D., N.D., C.V.D., S.E., R.G., R.J.G., M. Guckenberger, M. Götz, S.M.H., P.L., S. Leger, R.T.H.L., K.H.M.H., O.M., H.M., S.N., C.N., F.O., S.P., A.R., J.S., M.M.S., J.S.F., E.S., R.J.H.M.S., D.T., T.U., V.V., L.V.v.D., J.v.G., F.H.P.v.V.; statistical analysis, A.Z., M.A.A., S.B., I.B., N.D., R.J.G., S.M.H., O.M., H.M., S.P., E.A.G.P., M.M.S., J.S.F., R.J.H.M.S., D.T., V.V.; and manuscript editing, A.Z., M.V., M.A.A., H.J.W.L.A., V.A., A.A., S.B., R.J.B., R.B., M.B., L.B., I.B., G.J.R.C., C.D., A.D., N.D., I.E.N., A.Y.F., R.J.G., V.G., M. Guckenberger, M. Götz, M.H., P.L., S. Leger, R.T.H.L., K.H.M.H., O.M., H.M., S.N., S.P., E.A.G.P., A.U.K.R., M.M.S., N.M.S., J.S.F., E.S., R.J.H.M.S., D.T., E.G.C.T., V.V., L.V.v.D., J.v.G., F.H.P.v.V., C.R., S. Löck

The authors received funding from the Cancer Research UK and Engineering and Physical Sciences Research Council, with the Medical Research Council and the Department of Health and Social Care (C1519/A16463: M.M.S., G.C., V.G.), Dutch Cancer Society (10034: R.B.), EU Seventh Framework Programme (ARTFORCE 257144: R.T.H.L., P.L.; REQUITE 601826: R.T.H.L., P.L.), Engineering and Physical Sciences Research Council (EP/M507842/1: P.W., E.S.; EP/N509449/1: P.W., E.S.), European Research Council (ERC AdG-2015: 694812-Hypoximmuno: R.T.H.L., P.L.; ERC StG-2013: 335367 bioIRT: D.T.), Eurostars (DART 10116: R.T.H.L., P.L.; DECIDE 11541: R.T.H.L., P.L.), French National Institute of Cancer (C14020NS: M.C.D., M.H.), French National Research Agency (ANR-10-LABX-07-01: M.C.D., M.H.; ANR-11-IDEX-0003-02: C.N., F.O., I.B.), German Federal Ministry of Education and Research (BMBF-03Z1N52: A.Z., S. Leger, E.G.C.T., C.R.), Horizon 2020 Framework Programme (BD2Decide PHC-30-689715: R.T.H.L., P.L.; IMMUNOSABR SC1-PM-733008: R.T.H.L., P.L.), Innovative Medicines Initiative (IMI JU QuIC-ConCePT 115151: R.T.H.L., P.L.), Interreg V-A Euregio Meuse-Rhine (Euradiomics: R.T.H.L., P.L.), National Cancer Institute (P30CA008748: A.A.; U01CA187947: S.E., S.N.; U24CA189523: S.B., S.P., S.M.H., C.D.), National Institute of Neurologic Disorders and Stroke (R01NS042645: S.B., S.P., S.M.H., C.D.), National Institutes of Health (R01CA198121: A.A.; U01CA143062: R.J.G.; U01CA190234: J.v.G., A.Y.F., H.J.W.L.A.; U24CA180918: A.Y.F.; U24CA194354: J.v.G., A.Y.F., H.J.W.L.A.), SME phase 2 (RAIL 673780: R.T.H.L., P.L.), Swiss National Science Foundation (310030 173303: M.B., S.T.L., M. Guckenberger; PZ00P2 154891: A.D.), Technology Foundation STW (10696 DuCAT: R.T.H.L., P.L.; P14-19 Radiomics STRaTegy: R.T.H.L., P.L.), the Netherlands Organization for Health Research and Development (10-10400-98-14002: R.B.), the Netherlands Organization for Scientific Research (14929: E.A.G.P., R.B.), University of Zurich Clinical Research Priority Program (Tumor Oxygenation: M.B., S.T.L., M. Guckenberger), and the Wellcome Trust (WT203148/Z/16/Z: M.M.S., G.C., V.G.).

Disclosures of Conflicts of Interest: A.Z. disclosed no relevant relationships. M.V. disclosed no relevant relationships. M.A.A. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: works for H. Lee Moffitt Cancer Center and Research Institute. Other relationships: disclosed no relevant relationships. H.J.W.L.A. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: receives payment for board membership from Sphera and Genospace; holds stock/stock options in Sphera and Genospace. Other relationships: disclosed no relevant relationships. V.A. disclosed no relevant relationships. A.A. disclosed no relevant relationships. S.A. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: works for Siemens Healthineers. Other relationships: disclosed no relevant relationships. S.B. disclosed no relevant relationships.

R.J.B. disclosed no relevant relationships. **R.B.** disclosed no relevant relationships. **M.B.** Activities related to the present article: author's institution received a grant from the Department of Radiation Oncology at University Hospital Zurich. Activities not related to the present article: has grants/grants pending with the Department of Radiation Oncology at University Hospital Zurich. Other relationships: disclosed no relevant relationships. **L.B.** disclosed no relevant relationships. **I.B.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: works for the French National Center for Scientific Research; received money for a pending European patent; has received reimbursement for travel/accommodations/meeting expenses unrelated to activities listed. Other relationships: disclosed no relevant relationships. **G.J.R.C.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for NanoMab Technology; author's institution has grants/grants pending with NanoMab Technology and Theragnostics. Other relationships: disclosed no relevant relationships. **C.D.** disclosed no relevant relationships. **A.D.** disclosed no relevant relationships. **M.C.D.** disclosed no relevant relationships. **N.D.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has received payment for teaching courses for the European Society for Radiotherapy and Oncology. Other relationships: disclosed no relevant relationships. **C.V.D.** disclosed no relevant relationships. **S.E.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: works for Listo Unlimited. Other relationships: disclosed no relevant relationships. **I.E.N.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has patents planned, pending or issued with the University of Michigan. Other relationships: disclosed no relevant relationships. **A.Y.F.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has received payment from educational institutions for lecture honorariums; received a contract award to develop Imaging Data Commons for the National Institutes of Health/National Cancer Institute. Other relationships: disclosed no relevant relationships. **R.G.** disclosed no relevant relationships. **R.J.G.** Activities related to the present article: author's institution received payment from HealthMyne for the provision of writing assistance, medicines, equipment, or administrative support. Activities not related to the present article: is a consultant for HealthMyne; has patents submitted and issued with the H. Lee Moffitt Cancer Center; holds stock/stock options in HealthMyne. Other relationships: disclosed no relevant relationships. **V.G.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received reimbursement for travel/accommodations/meeting expenses from Siemens Healthcare. Other relationships: disclosed no relevant relationships. **M. Götz.** disclosed no relevant relationships. **M. Guckenberger.** disclosed no relevant relationships. **S.M.H.** disclosed no relevant relationships. **M.H.** disclosed no relevant relationships. **E.I.** disclosed no relevant relationships. **P.L.** Activities related to the present article: author's institution received a grant from Netherlands Wetenschappelijk Organisatie. Activities not related to the present article: receives payment for board membership from OncoRadiomics; has grants/grants pending with European commissions; has patents planned, pending or issued with Maastricht University; holds stock/stock options in OncoRadiomics. Other relationships: institution has radiomics patents licensed to OncoRadiomics; institution receives royalties related to radiomics patents licensed to OncoRadiomics. **S. Leger.** disclosed no relevant relationships. **R.T.H.L.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: works for OncoRadiomics; holds stock/stock options in OncoRadiomics. Other relationships: receives money for patents pending and issued. **J.L.** disclosed no relevant relationships. **E.L.** disclosed no relevant relationships. **A.L.** disclosed no relevant relationships. **K.H.M.H.** disclosed no relevant relationships. **O.M.** disclosed no relevant relationships. **H.M.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: works for Haute Ecole Spécialisée de Suisse occidentale. Other relationships: disclosed no relevant relationships. **S.N.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: receives payment for board membership from Fovia, EchoPixel, and Radlogics; has consulted for Carestream; has patents planned, pending or issued with the U.S. Patent and Trademark Office. Other relationships: disclosed no relevant relationships. **C.N.** disclosed no relevant relationships. **F.O.** disclosed no relevant relationships. **S.P.** disclosed no relevant relationships. **E.A.G.P.** Activities related to the present article: author's institution received a grant from STRaTeGy. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **A.R.** disclosed no relevant relationships. **A.U.K.R.** Activities related to the present article: is a consultant for Voxal Analytics; received support for travel to meetings for study or other purposes from Cambridge Healthtech Institute. Activities not related to the present article: is a consultant for high-content screening with Voxal Analytics; has grants/grants pending with federal agencies; has filed disclosures with MD Anderson Cancer Center for radiomics. Other relationships: disclosed no relevant relationships. **J.S.** disclosed no relevant relationships. **M.M.S.** disclosed no relevant relationships. **N.M.S.** disclosed no relevant relationships. **J.S.F.** disclosed no relevant relationships. **E.S.** disclosed no relevant relationships. **R.J.H.M.S.** disclosed no relevant relationships. **S.T.L.** disclosed no relevant relationships. **D.T.** Activities related to the present article: disclosed no

relevant relationships. Activities not related to the present article: has grants/grants pending with Elekta and Siemens Healthineers; has received payment for lectures, including service on speakers bureaus, from Elekta and Siemens Healthineers. Other relationships: disclosed no relevant relationships. **E.G.C.T.** disclosed no relevant relationships. **T.U.** disclosed no relevant relationships. **V.V.** Activities related to the present article: author's institution received a grant from Varian, Elekta, Merck Serono, and ViewRay; author's institution received a consulting fee or honorarium from Beta Glue; author's institution has received support for travel to meetings for study or other purposes from Merck Serono. Activities not related to the present article: disclosed no relevant relationships. Other relationships: institution receives money for a patent on a brachytherapy MRI device. **L.V.v.D.** disclosed no relevant relationships. **J.v.G.** disclosed no relevant relationships. **E.H.P.v.V.** disclosed no relevant relationships. **P.W.** Activities related to the present article: author's institution received a grant from the Engineering and Physical Sciences Research Council-Doctoral Training Partnership. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **C.R.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received individual funding as a lecturer from Siemens Healthineers. Other relationships: disclosed no relevant relationships. **S. Löck.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is an editor for *Radiotherapy and Oncology* published by Elsevier. Other relationships: disclosed no relevant relationships.

References

1. La Thangue NB, Kerr DJ. Predictive biomarkers: A paradigm shift towards personalized cancer medicine. *Nat Rev Clin Oncol* 2011;8(10):587–596.
2. O'Connor JPB, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 2017;14(3):169–186.
3. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14(12):749–762.
4. Morin O, Vallières M, Jochems A, et al. A Deep Look into the Future of Quantitative Imaging in Oncology: A Statement of Working Principles and Proposal for Change. *Int J Radiat Oncol Biol Phys* 2018;102(4):1074–1082.
5. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016;278(2):563–577.
6. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5(1):4006.
7. Sun R, Limkin EJ, Vakalopoulou M, et al. A radiomics approach to assess tumor-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: An imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018;19(9):1180–1191.
8. Lu H, Arshad M, Thornton A, et al. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer. *Nat Commun* 2019;10(1):764.
9. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: Bridging imaging and genomics. *Abdom Radiol (NY)* 2019;44(6):1960–1984.
10. Leandrou S, Petroudi S, Kyriacou PA, Reyes-Aldasoro CC, Pattichis CS. Quantitative MRI Brain Studies in Mild Cognitive Impairment and Alzheimer's Disease: A Methodological Review. *IEEE Rev Biomed Eng* 2018;11:97–111.
11. Chaddad A, Desrosiers C, Toews M. Multi-scale radiomic analysis of sub-cortical regions in MRI related to autism, gender and age. *Sci Rep* 2017;7(1):45639.
12. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, et al. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology* 2018;288(2):407–415.
13. Welch ML, McIntosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiother Oncol* 2019;130:2–9.
14. Meyer M, Ronald J, Vernuccio F, et al. Reproducibility of CT Radiomic Features within the Same Patient: Influence of Radiation Dose and CT Reconstruction Settings. *Radiology* 2019;293(3):583–591.
15. Kalpathy-Cramer J, Mamomov A, Zhao B, et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. *Tomography* 2016;2(4):430–437.
16. Bogowicz M, Leijenaar RTH, Tanadini-Lang S, et al. Post-radiochemotherapy PET radiomics in head and neck cancer: The influence of radiomics implementation on the reproducibility of local control tumor models. *Radiother Oncol* 2017;125(3):385–391.
17. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: The past, the present... any future? *Eur J Nucl Med Mol Imaging* 2017;44(1):151–165.

18. Foy JJ, Robinson KR, Li H, Giger ML, Al-Hallaq H, Armato SG 3rd. Variation in algorithm implementation across radiomics software. *J Med Imaging (Bellingham)* 2018;5(4):044505.
19. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol Biol Phys* 2018;102(4):1143–1158.
20. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015;60(14):5471–5496.
21. Lambin P. Radiomics Digital Phantom. <https://doi.org/10.17195/candata.2016.08.1>. Published 2016. Accessed April 24, 2019.
22. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging* 2013;26(6):1045–1057 <https://doi.org/10.1007/s10278-013-9622-7>.
23. Vallières M, Freeman CR, Skamene SR, El Naqa I. Data from: A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. The Cancer Imaging Archive. Published 2015. Accessed October 18, 2019.
24. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014;67(4):401–409.
25. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979;86(2):420–428.
26. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1(1):30–46.
27. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2):155–163.
28. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: A systematic review on artificial intelligence and radiomics. *Eur J Nucl Med Mol Imaging* 2019;46(13):2656–2672.
29. Orlhac F, Soussan M, Maisonneuve JA, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: Relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med* 2014;55(3):414–422.
30. van Timmeren JE, Leijenaar RTH, van Elmpt W, et al. Test-Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific? *Tomography* 2016;2(4):361–365.
31. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD Statement. *Br J Surg* 2015;102(3):148–158.
32. Depeursinge A, Foncubierta-Rodriguez A, Van De Ville D, Müller H. Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Med Image Anal* 2014;18(1):176–196.
33. Zwanenburg A. Radiomics in nuclear medicine: Robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *Eur J Nucl Med Mol Imaging* 2019;46(13):2638–2655.
34. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a Method to Compensate Multicenter Effects Affecting CT Radiomics. *Radiology* 2019;291(1):53–59.
35. Choe J, Lee SM, Do KH, et al. Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. *Radiology* 2019;292(2):365–373.